DOCUMENT RESUME

ED 457 860 IR 058 318

AUTHOR	Smith, Abby
TITLE	Strategies for Building Digitized Collections. Strategies and Tools for the Digital Library.
INSTITUTION	Council on Library and Information Resources, Washington, DC. Digital Library Federation.
ISBN	ISBN-1-887334-87-4
PUB DATE	2001-00-00
NOTE	44p.
AVAILABLE FROM	Digital Library Federation, Council on Library and Information Resources, 1755 Massachusetts Ave., NW, Suite 500, Washington, DC 20036 ($15). Tel: 202-939-4750; Fax: 202-939-4765; e-mail: info@clir.org; Web site: http://www.clir.org/diglib. For full text: http://www.clir.org/pubs/reports/pub101/contents.html.
PUB TYPE	Reports - Evaluative (142)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	Access to Information; Information Technology; *Library Collection Development; Library Materials; Online Systems; *Preservation; Program Development; *Research Libraries
IDENTIFIERS	*Digital Collections; *Digital Technology; Digitizing; Rare Books; Special Collections (Library)

ABSTRACT

In April 2000, the Digital Library Federation commissioned three reports to address broader concerns about digital collections in research libraries. This report synthesizes the nearly 10 years' experience that libraries have had digitizing items from their rare, special, and general collections, and making them available online. The report demonstrates that digitization programs work best where their role within a library's collection development strategy is clearly understood, and identifies several roles that such programs can play. The author muses about the extent to which digitally reformatted special and rare collections can actually support scholarly research, and looks at whether leading research libraries in particular might more usefully focus on digitizing general as opposed to special and rare collections. The report opens with points to consider in developing a sustainable strategy. The second section addresses identification, evaluation and selection, discussing polices, guidelines and best practices, and rationales for digitization. The third section focuses on institutional impacts and discusses treatment and disposition of source materials, scalability, intellectual control and data management, coordinated collection development, funding, preservation, and support of users. A final section addresses challenges in evaluating costs and benefits, and offers recommendations. (Contains 47 references.) (AEF)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 457 860

Strategies for Building Digitized Collections

by Abby Smith

September 2001

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

B.H. Leneyr

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

IR058318

Digital Library Federation
and
Council on Library and Information Resources

2

BEST COPY AVAILABLE

This report is one in a series from CLIR on
STRATEGIES AND TOOLS FOR THE DIGITAL LIBRARY

Strategies for Building Digitized Collections

by Abby Smith

September 2001

Digital Library Federation
Council on Library and Information Resources
Washington, D.C.

About the Author

Abby Smith is director of programs at the Council on Library and Information Resources (CLIR). She is responsible for developing and managing collaboration with key library and archival institutions to ensure long-term access to our cultural and scholarly heritage. Before joining CLIR in 1997, she had worked at the Library of Congress for nine years, first as a consultant to the special collections research divisions, then as coordinator of several cultural and academic programs in the offices of the Librarian of Congress and the Associate Librarian for Library Services. She directed a preservation microfilming program in the former Soviet Union, curated three exhibitions of Russian library and archival treasures from the former Soviet Union, and was curator and project director for the library's first-ever permanent exhibition of its holdings, *Treasures of the Library of Congress*.

ISBN 1-887334-87-4

Published by:

Digital Library Federation
Council on Library and Information Resources
1755 Massachusetts Avenue, NW, Suite 500
Washington, DC 20036

Web site at http://www.clir.org

Additional copies are available for $20 per copy. Orders must be placed online through CLIR's Web site.

The paper in this publication meets the minimum requirements of the American National Standard for Information Sciences—Permanence of Paper for Printed Library Materials ANSI Z39.48-1984.

Copyright 2001 by the Council on Library and Information Resources. No part of this publication may be reproduced or transcribed in any form without permission of the publisher. Requests for reproduction should be submitted to the Director of Communications at the Council on Library and Information Resources.

Contents

Preface iv

Acknowledgments vi

1. Introduction 1
    1.1. Defining a Sustainable Strategy 2

2. Identification, Evaluation, and Selection 4
    2.1. Policies, Guidelines, and Best Practices 4
    2.2. Rationales for Digitization 7
        2.2.1. Preservation 7
            2.2.1.1. Surrogates 7
            2.2.1.2. Replacements 7
        2.2.2. Access 8
            2.2.2.1. Access for Outreach and Community Goals 8
            2.2.2.2. Building Collections 9
            2.2.2.3. Meeting Use and Infrastructure Needs 10
                General Collections 12
                Special Collections 13

3. Institutional Impacts 15
    3.1. Treatment and Disposition of Source Materials 15
    3.2. Scalability 16
    3.3. Intellectual Control and Data Management 19
    3.4. Coordinated Collection Development 21
    3.5. Funding 22
    3.6. Preservation 23
    3.7. Support of Users 25

4. Conclusions and Recommendations 26
    4.1. Costs 26
    4.2. Benefits 28
    4.3. Looking Ahead 29
    4.4. Recommendations 30

5. References 32

Preface

In January 2000, the Digital Library Federation (DLF) launched an informal survey to identify the major challenges confronting research libraries that use information technologies to fulfill their curatorial, scholarly, and cultural missions. With astonishing unanimity of opinion and clarity of voice, respondents pointed to digital collection development as their single greatest challenge. Whether the digital information came from a commercial publisher or from a digitization unit within the library, it seemed to exist under a cloud of profound and unsettling uncertainty. Would it be useful and useable in its present or intended form, or require additional work by catalogers, systems staff, or subject bibliographers? What new demands would its availability make on library reference staff? What level of continued investment would be necessary to ensure its accessibility on current hardware and software?

The survey also revealed that leading research libraries had learned a great deal about their digital collections through experience. Though substantial, that learning had rarely been expressed outside the collection policies, working papers, and implementation guidelines that libraries create to coordinate and manage their collection development efforts. Accordingly, in April 2000, the DLF commissioned three reports to address broader concerns about digital collections. They are: *Building Sustainable Collections of Free Third-Party Web Resources*, by Louis Pitschmann, *Selection and Presentation of Commercially Available Electronic Resources: Issues and Practices*, by Timothy Jewell, and the report before you. The reports mark a starting point for what we hope will emerge as an evolving publication series.

Working to a common outline and based on the lessons of experience, the authors demonstrate how decisions taken by a library when acquiring (or creating) electronic information influence how, at what cost, and by whom the information will be used, maintained, and supported. By assembling and reviewing current practice, the reports aim where possible to document effective practices. In most cases, they are able at least to articulate the strategic questions that libraries will want to address when planning their digital collections.

In this report, Abby Smith synthesizes the nearly 10 years' experience that libraries have had digitizing items from their rare, special, and general collections, and making them available online. The learning she uncovers is distilled in and extended by several case studies conducted in leading digital libraries with very different digitization programs. Smith demonstrates that digitization programs work best where their role within a library's collection development strategy is clearly understood, and she identifies several roles

that such programs can play. Smith also asks a number of searching questions. She muses about the extent to which digitally reformatted special and rare collections can actually support scholarly research. Probing further, she wonders whether leading research libraries in particular might more usefully focus on digitizing general as opposed to special and rare collections. In this way, they would make important holdings available in new ways while taking a first step in avoiding costs associated with their redundant management. The report is consequently much more than a strategic guide for individual institutions; it is a route map that points important directions for the library community as a whole.

Daniel Greenstein
Director, Digital Library Federation

Acknowledgments

This report would not have been possible without the cooperation of the six libraries that opened their doors to me, invited my direct questions, and answered them frankly and with great thoughtfulness. I thank the library directors and their staff at Cornell University, the New-York Historical Society, The New York Public Library, New York University, the University of Michigan, and the University of Virginia. I profited from their time, attention, and careful review of a preliminary draft for accuracy. Others who have read this report in various drafts and offered their suggestions for improvement include Ross Atkinson, Ricky Erway, Dale Flecker, Janet Gertz, Peter Hirtle, Anne Kenney, and Bonnie MacEwan. I worked closely with Daniel Greenstein, Tim Jewell, and Lou Pitschmann in developing a framework for this report and reviewing each draft. My special thanks go to Kathlin Smith, the ideal editor.

1. Introduction

Libraries have been digitizing collections for a decade or more. Their collective experience has produced a depth of technical expertise and a set of tested practices. That information is widely shared among digital library staffs and has been well reported at meetings and in publications. This ongoing experiment with representing research collections online has resulted in the codification of technical practices and the emergence of clear trends in selection policies. This paper reviews existing selection practices in libraries, identifies selection policies and best practices where they exist, and discusses the long-term implications of the opportunities and constraints that shape digital-conversion programs. This is not a systematic review of what all research libraries are doing, but an analysis of significant achievements that will make it possible to identify good practices and benchmarks for success. Every library, regardless of size or mission, will need to determine for itself how and when digitization will move from being an experiment to becoming a collection-development strategy that is well integrated into its daily practice.

For purposes of analysis, this study looks primarily at a subset of "first-generation" digital libraries, that is, those that have been engaged in significant digitization projects for a while. However, the study also looks at a few libraries that are just beginning to develop digitization programs to see what approaches they have taken, in light of others' experience. Research was conducted by studying the Web sites of all Digital Library Federation (DLF) members as well as the sites of other libraries and research institutions engaged in putting collections online. More important for analytical purposes were the site visits made to selected libraries—the University of Michigan, Cornell University, the University of Virginia (UVA), The New York Public Library (NYPL), New York University (NYU), and the New-York Historical Society. In addition to the fact that some of the selected institutions are first-generation digital libraries and others are not, there are great differences in governance and funding among the libraries surveyed. Some are in public universities, some are in private institutions, and some are independent of an academic institution.

These differences are reflected in the various approaches they take to selecting what to digitize, how to do so, and for whom.

Each library was given a set of questions about selection criteria that constituted the framework for investigation, and each institution organized its responses individually. (The Library of Congress [LC], also included in this study, answered the questions in writing, and no site visit was made.) The questions begin with the selection process and proceed through the creation of metadata, decisions about access policies, and user support systems.

1.1. Defining a Sustainable Strategy

While the great majority of research libraries have undertaken digitization projects of one type or another, only a few are developing full-scale digitization programs rather than focusing on discrete uses of digitization for specific purposes. How do the libraries that have undertaken full-scale efforts conceptualize the role of digitized collections in providing collections and services to their core constituencies? What are they doing, or what have they determined must be done, to move from project-based conversion to programs that, whether large or small, have a well-defined role in the long-term goals of the library?

This report works from the assumption that to be sustainable, a digitization program should have certain intrinsic features. It should

- be integrated into the fabric of library services;
- be focused primarily on achieving mission-related objectives;
- be funded from predictable streams of allocation, be they external or internal; and
- include a plan for the long-term maintenance of its assets.

A sustainable digitization program, in other words, would be fully integrated into a library's traditional collection-development strategies. A digitization program need not be large and production oriented to be sustainable. The role of conversion can be significant and well thought out, even when the conversion program serves limited purposes and has limited resources.

Any assessment of what libraries have achieved so far must take into account two key factors common to sustainable collection development, be it of analog, digitized, or born-digital materials. These factors are

- a *strategic view* of the role of collections in the service of research and teaching or other core institutional missions, and
- *life cycle planning* for the collections, beginning with their identification and including acquisition, cataloging and preservation, and providing reference.

A strategic view can be revealed in many cases not only by looking at how closely the results serve the mission but also at the decision-making process itself—that is, who decides what to convert to serve which ends. When are the decisions made primarily by subject

specialists based on existing collection strengths, and when is the selection process shaped by curricular development and other faculty needs? If the latter, then by what process are the faculty involved and how are teaching and research tools developed to meet their needs?

Ensuring long-term access to digital collections depends on careful life-cycle management. How does the library budget for not only the creation of the digital scans but also for the metadata, storage capacity, preservation tools (e.g., refreshing, migration), and user support—the sorts of things that are routinely budgeted for book acquisitions? How much of the program is supported by grant funding and how much by base funding? If the program is currently grant supported, what plans exist to make it self-sustaining? A sustainable digitization strategy may well include the creation of digital surrogates that serve short-term needs and do not demand long-term support. The crucial thing is to anticipate what support, if any, will be needed.

Selecting materials for digitization is more complex than is selecting materials for the purchase or licensing of born-digital materials, because it involves expending resources for items that are already in the library's collection rather than acquiring new ones. In theory, a library would choose to digitize existing collection items only if it could identify the value that is added by digitization and determine that the benefits outweigh the costs. But in practice, the research library community has, over the past decade, gone boldly forth with digitization projects not knowing how to measure their costs or benefits. Digitization technology and its costs are constantly changing; as a result, budgeting models that make comparisons between libraries can be meaningless or downright misleading. Unlike selecting officials who decide the purchase or license of electronic resources, those responsible for digital conversion do not have a set of fixed prices for services and collections on offer. The only way for many libraries to get at the issue of cost is to undertake projects for their own sake, in the expectation that documentation of expenditures will yield some meaningful data. Libraries that have been able to secure funding for projects, document their activities and expenditures, and share that information with their colleagues have emerged as the leaders of the community, if only because of their policies to share their knowledge. Their experiences are more relevant for this report than are those of others who have embarked on fewer projects or who have failed to document and share their knowledge.

The other unknown factor in this first decade has been the benefit—the potential of this technology to enhance teaching, research, lifelong learning, or any number of possible goals that digitization is intended to achieve. How could we know in advance how users other than ourselves would adapt this technology? How could we conceptualize use of digitally reborn collections except by extrapolating what we know from the analog realm? Regrettably, most academic institutions, despite their clearly stated goals of improving or at least enhancing research and teaching, have done less than they might

have to gather meaningful data about the uses of digitized collections. While this report does address issues of costs and benefits, it should be remembered that as a community, we still have insufficient data on which to draw firm conclusions and base recommended practices.

This report aims to synthesize experiences in order to identify trends, accomplishments, and problems common to libraries and to many cultural institutions when they represent their collections online. A brief review of rationales for digitization is followed by a discussion of the ways in which digitization affects an institution. What consequences, intended or not, result from selection decisions? What factors, such as funding, may set constraints on decision making? One of the chief factors influencing selection decisions is copyright. This topic will not be explored here in detail because of its complexity, but considerations related to rights management are often foremost in mind when librarians assess collections for digitization. At the program planning stages, copyright is often viewed from the point of view of risk management.[1] Eliminating materials that are or might be under copyright reduces the risk of infringement to zero. But how does that affect the concepts of completeness and fair use that traditionally guide library access policies?

2. Identification, Evaluation, and Selection

2.1. Policies, Guidelines, and Best Practices

A great deal has been written on the subject of selection for digitization and on the management of conversion projects. Much of this literature is published on the Web and has become de facto "best practice," to the extent that many institutions applying for digitization grants use it to plan their projects and develop selection criteria. In addition to these guidelines, there are a number of reports about selection for digitization that range from project management handbooks and technical guides to imaging, to broad, nontechnical articles aimed at those outside the library community who fund such programs.[2]

Very few libraries have developed their own formal written policies for conversion criteria. Those that do have such documents tend to refer to them as "guidelines." These documents tend to focus on technical aspects of selection and, even more, on project planning. When asked why they do not have a policy, most institutions reply that it is too early to formulate policies, that they have not gotten around to formulating them, or that the institution does not have

[1] Literature on copyright abounds; among the most useful in program planning is by Melissa Levine in Sitts 2000.

[2] See Research Libraries Group 1996; Digital Library Federation and Research Libraries Group 2000; Sitts 2000; Smith 1999; Gertz 2000; de Stefano 2001; Kenney and Rieger, 2000; Library of Congress National Digital Library Program 1997.

written collection-development policies for other materials so it is unlikely to write them for digitized collection development.

These documents almost always focus on the planning of digital projects or of various elements of a larger program, rather than on the rationale for digitization. The University of Michigan, for example, has a written policy that clearly aims to fit digitization into the context of traditional collection development. It states that "core questions" underlying digitization should be familiar to any research library collection specialist (University of Michigan 1999). These questions are as follows:

- Is the content original and of substantial intellectual quality?
- Is it useful in the short and/or long term for research and instruction?
- Does it match campus programmatic priorities and library collecting interests?
- Is the cost in line with the anticipated value?
- Does the format match the research styles of anticipated users?
- Does it advance the development of a meaningful organic collection?

These are fundamental collection-development criteria that assert the importance of the research value of source materials over technical considerations; however, they are quite general. The rest of Michigan's policy focuses not on how to select items for conversion but on how anticipated use of the digital surrogates should affect decisions about technical aspects of the conversion, markup, and presentation online.

Harvard's selection criteria offer far more detailed considerations than do those of Michigan (Hazen, Horrell, and Merrill-Oldham 1998). In common with the Michigan criteria, the Harvard criteria focus largely on questions that come *after* the larger, "why-bother-to-digitize-this-rather-than-that" issues have already been answered. Creation of digital surrogates for preservation purposes is cited as one legitimate reason for selection, as are a number of considerations aimed not at preservation but solely at increasing access. (Sometimes digitization does both at once, as in the case of rare books or manuscripts.)

The Harvard guidelines have been useful to many beyond Harvard who are engaged in planning conversion projects because they present a matrix of decisions that face selectors and are available on the Web (Brancolini 2000). The authors begin with the issue of copyright—whether or not the library has the right to reformat items and distribute them in limited or unlimited forms. They then ask a series of questions derived from essentially two points of departure:

Source material: Does it have sufficient intellectual value to warrant the costs? Can it withstand the scanning process? Would digitization be likely to increase its use? Would the potential to link to other digitized sources create a deeper intellectual resource? Would the materials be easier to use?

Audience: Who is the potential audience? How are they likely to use the surrogates? What metadata should be created to enhance use?

The answers to these and similar questions should guide nearly all the technical questions related to scanning technique, navigational tools and networking potential, preservation strategy, and user support.

The primary nontechnical criterion—research value—is a subjective one and relies on many contingencies for interpretation. What does it mean to say that something has intrinsic research value? Do research libraries collect any items that do not have such value? Should we give priority to items that have research value today or to those that may have it tomorrow? What relationship does current demand have to intrinsic value? Because the answers to these questions are subjective, the only things excluded under these selection criteria are items that are difficult to scan (for example, oversized maps) or things that are very boring or out of intellectual fashion. Interestingly, foreign language materials are nearly always excluded from consideration, even if they are of high research value, because of the limitations of optical character recognition (OCR) software and because they often have a limited number of users. There are digital projects that have converted valuable historical sources, from Egyptian papyri to medieval manuscripts, into image files (such as the Advanced Papyrological Information System [APIS] and the Digital Scriptorium). In general, however, the conversion of non-English language sources into searchable text continues to be rare.

This high-level criterion of research value is also an intrinsic part of traditional collection-development policies. The difference is that in most libraries, the acquisition of monographs, to take an example, fits into a longstanding activity that has been well defined by prior practice. This practice governs the acquisition of new materials, that is, those that the library does not already hold. (The issue of how many copies is secondary to the decision to acquire the title.) Selection of an item for digitization is reselection, and the criteria for its digitization, or repurposing, will be different from those for its acquisition. The meaning of research value will also differ, because the methods of research used for digital materials differ from those used for analog, and the types of materials that are mined—and how—are also fundamentally different. Several large digitization programs today are grounded in the belief that it is the nature of research itself that is "repurposed" by this technology, and it is often surprising to see which source material yields the greatest return when digitized.

As one librarian said, the guidelines addressing selection that are used routinely, whether official or not, are by and large "project oriented." It would be a mistake to confuse what libraries are doing now with what libraries should and would do if "we understood what higher purpose digitization serves." While guidelines for technical matters such as image capture and legal rights management are extremely useful and should be codified, formal collection-development policies are still a long way off.

2.2. Rationales for Digitization

Libraries usually identify two reasons for digitization: to preserve analog collections, and to extend the reach of those collections. Most individual projects and full-scale programs serve a mix of both purposes. As librarians have learned from tackling the brittle book problem through deacidification and reformatting, it is difficult and often pointless to separate preservation and access. When a library is seeking outside funding for digital conversion (apparently still the primary source of funding for many libraries), it tends to cite as many possible benefits from conversion as possible. For this reason, preservation and access are usually mentioned in the same breath. Nonetheless, because it has been generally conceded that digital conversion is not as reliable for preservation purposes as is microfilm reformatting, it is worthwhile to consider what institutions are doing in terms of preservation per se.

2.2.1. Preservation

2.2.1.1. Surrogates

The use of scans made of rare, fragile, and unique materials—from prints and photographs to recorded sound and moving image—is universally acclaimed as an effective tool of preventive preservation. For materials that cannot withstand frequent handling or, because of their value or content, pose security risks, digitization has proved to be a boon.

2.2.1.2. Replacements

For paper-based items, librarians generally agree that digital scans are the preferred type of preservation surrogates. They are widely embraced by scholars and are preferred over microfilm. However, most librarians also assert that scanning *in lieu of filming* does not serve preservation purposes, because the expectation that we can migrate those scans into the future is simply not as great as is our conviction that we can manage preservation microfilm over decades. There is a general hope that the problem of digital longevity will soon be resolved. In anticipation of that day, most libraries are creating "preservation-quality" digital masters—scans that are rich enough to use for several different purposes and are created to obviate the need to rescan the original. These masters are sometimes created together with preservation master microfilm.

Only one institution, the University of Michigan, has a policy to scan brittle books and use the scans as replacements rather than as surrogates. The university has created a policy for the selection and treatment of these books, and it explicitly talks of digital replacements as a crucial strategy for collection management (University of Michigan 1999). This policy is based on the premise that books printed on acid paper have a limited life span and that, for those items with insignificant artifactual value, the library is not only rescuing the imperiled information but also making it more accessible by scanning in lieu of filming. (The preservation staff continues to microfilm items identified by selectors for filming as well as to deacidi-

fy volumes that are at risk but not yet embrittled.) The focus of Michigan's digital program is the printed record, not special collections such as rare books and photographs, and digitization has been made a key collection-management tool for these holdings. Cornell also has incorporated digitization into collection management (that is, it is not for access alone), although its efforts are not as systematic as are those of Michigan. At Cornell there is a preference for digital replacements of brittle materials with backup to computer output microfilm (COM) or replacement hard copies made from digital scans. The Library of Congress has also begun implementing preservation strategies based on digitization in its project to digitize the nineteenth-century journal, *Garden and Forest*. Most libraries, though, elide the issue of digitally replacing brittle materials because they scan chiefly items from special collections.

For audiovisual materials, digital replacements appear to be inevitable, although standards for archival-quality re-recording have yet to be established. Because the recording media used for sound and moving image demand regular, frequent, and ultimately destructive reformatting, migrating onto digital media for preservation, as well as access, is acknowledged to be the only course to pursue for long-term maintenance. The University of California at Los Angeles (UCLA) and LC, both deeply engaged in audiovisual preservation, intend to digitize analog materials to provide long-term access. This does not mean that these institutions will dispose of the original analog source materials, only that the preservation strategy for these items will not be based on routine use of that analog source material.

2.2.2. Access

In nearly all research libraries, digitization is viewed as service of collections in another guise—one that provides enhanced functionality, convenience, some measure of preservation, aggregation of collections that are physically dispersed, and greatly expanded reach. Among all the strands of digitization activities at major research institutions, there are essentially three models of collection development based on access: one that serves as outreach to various communities; one that is designed to build collections; and one that is driven by a specific need, such as demand by a user or preservation surrogates, or that is part of a larger effort to develop core infrastructure. All libraries engage in the first kind of access to one degree or another. Significant strategic differences are evident, however, in their approaches to the choice between mounting large bodies of materials in the expectation of use versus collaborating with identified users to facilitate their data creation.

2.2.2.1. Access for Outreach and Community Goals

There will continue to be times when academic libraries create digital surrogates of their analog holdings for reasons that are important to the home institution yet not directly related to teaching and research. Libraries will continue to be parts of larger communities that look to them for purposes that transcend the educational mission of

the library per se. As custodians of invaluable institutional intellectual and cultural assets, libraries will always play crucial roles in fund raising, cultivating alumni allegiance, and public relations.

Occasions for selective digitization projects include exhibitions, anniversaries (when archives or annual reports often get into the queue), a funding appeal (digitization as a condition of donation), and efforts to build institutional identity. Careful consideration needs to be given to what goes online for whatever purpose because, once a collection is online, it becomes part of the institutional identity. Image building is a critical and often undervalued part of ensuring the survival of the library and its host institution. As custodians of the intellectual and cultural treasures of a university, libraries have an obligation to share that public good to the advantage of the institution.

2.2.2.2. *Building Collections*

The most common approach to digitization hinges on a collection-driven selection process in which a library decides to scan a set of materials identified by staff as having great research potential online. The terms "collection driven" and "use driven" are familiar from the preservation microfilming projects of the 1980s and 1990s. At that time, brittle books were conceptualized as a "national collection" that was held by a group of libraries, not just one. Each library could help preserve the national collection by filming a set of its holdings that were particularly strong, thereby avoiding duplication of effort by registering its filming activities and enhancing access to the endangered materials through the loan or duplication of microfilm. Librarians would select books not on the basis of their documented use (by checking circulation statistics or going with items that cross the circulation desk), but on the basis of whether they constituted a coherent set of monographs (and occasionally serials) arranged by subject matter or date of publication, or both.

This method, so well-known to preservation experts and subject specialists from the many grant-supported microfilming projects of the past two decades, has been transferred to the digital realm with one interesting twist: It has been extended primarily not to the general collections—monographs and serials—that are the heart of microfilming projects, but to special collections—materials that are rare or archival in nature or that are in nonprint format. This means that a library will scan items that exist as a defined collection, either by format (incunabula, daguerreotypes) or by genre of literature (antislavery pamphlets, travel literature gathered as a discrete group and held as such in the rare book department, photographs of the Reconstruction-era South given to the library by a donor and known under the donor's name, Sanborn fire insurance maps, and so forth).

Within each group, libraries may attempt to be as comprehensive as possible in putting items from the collection online to simulate the comprehensive or coherent nature of the source collection. Examples of such collection-driven digital collections are the Making of America (a subject and time period), Saganet (a set of special collection

items held by certain repositories that relate to Icelandic sagas), the Sam Nunn Papers project at Emory University, the Hoagy Carmichael site at Indiana University, and the Scenery Collection at the University of Minnesota.

In a survey of selection strategies at 25 research libraries, Paula de Stefano found that "the most popular approach to selecting collections for digital conversion is a subject-and-date parameter approach applied, by and large, to special collections, with little regard for use, faculty recommendations, scholarly input, editorial boards, or curriculum" (de Stefano 2001, 67). A recent analysis of 99 research libraries and their special collections done by the Association of Research Libraries (ARL) revealed that virtually all have been digitizing some of their special collections. The list of digitized collections submitted to ARL during this survey reveals just how eclectic these scanning projects are, a result fully consonant with de Stefano's findings about selection policies (Panitch 2001, 99, 116-123). This approach has often slyly been referred to as the "field of dreams" method of collection development ("build it and they will come"), implying a certain naïve hopefulness on the part of the selectors but also hinting at the elements of surprise and serendipity we see in the digital realm.

2.2.2.3. Meeting Use and Infrastructure Needs

Some libraries have decided that they will digitize collections only in response to explicit user-driven needs. As a state-supported institution, UVA has developed access projects that serve state and regional needs. These projects are based primarily on UVA's special collections holdings and are similar in scope and purpose to the type described above. The University of Virginia also has several digital conversion initiatives that are explicitly user driven, and these programs exist both in the library and elsewhere on campus. At the Institute for Advanced Technology in the Humanities (IATH), an academic center located in the library but administratively separate, scholars develop deep and deeply interpreted and edited digital objects that are, by any other name, publications. Examples include projects on the writers William Blake, Dante Gabriel Rossetti, and Mark Twain, as well as the Valley of the Shadow Civil War site. Within the library is the Electronic Text Center, where staff members choose to encode humanities texts that they put up without the interpretive apparatus of the IATH objects. These are analogous to traditional library materials that are made available for others to interpret; the difference is that encoded text is far more complicated a creature than is the OCR text that other libraries are creating. The Electronic Text Center is not so much responding to faculty or student demand as it is being driven by a technology. Exploring the potential of various encoding schemes is part of its agenda.

Under its Digital Library Initiative and with funds provided by the university, Harvard University libraries are concentrating on building an infrastructure to support born-digital materials first and foremost, rather than on building collections of digital surrogates of

existing collections. Where the libraries have converted items, the criteria for selection have to do with user needs, not general collection building. And while the holdings of the more than 100 repositories in the university certainly comprise a rich collection of cultural heritage, Harvard will attempt to serve the Harvard community, not the larger community (Flecker 2000). "While in many instances the digital conversion of retrospective materials already in the University's collections can increase accessibility and add functionality and value to existing scholarly resources, it is strategically much more important that the library begin to deal with the increasing flood of materials created and delivered solely in digital format." Although $5 million of the $12 million allocated by the university is for content development, so far the majority of content development comprises conversion-for-access purposes. Slated for review are the collections that have been mounted so far. "One specific issue being discussed is the randomness of the areas covered by the content projects. Since these depend upon the initiative of individuals, it is no surprise that the inventory of projects undertaken is spotty, and that there are notable gaps . . . . It is also possible that specific projects will be commissioned to address strategic topics" (Flecker 2000). However, the gaps Flecker refers to are not content per se—specific subjects that would complement one another—but content that demands different types of digital format—for example, encoded text, video, or sound recordings. This is a technical criterion, of course, independent of collection development, and is fully concordant with the purposes that Flecker identifies the initiative is to serve.

At New York University (NYU), the focus is on the user as part of a plan that allocates relatively modest resources for digitization. New York University presents collections of cultural significance through online exhibitions and other modes of Web outreach rather than engaging in full collection conversion. This library has decided to concentrate on working with faculty and graduate students to develop digital objects designed to enhance teaching and research through its Studio for Digital Projects and Research and its Faculty Technology Center. Like Harvard, NYU is giving priority to the development of an infrastructure to deal with born-digital materials and, in an institution with extensive programs and collections in the arts and performance studies, on multimedia archives converted to digital form for presentation. New York University plans to give grants to faculty members to develop teaching and research tools. However, the library staff is now putting much of its effort toward preparing for the time, seen to be imminent, when the demands of born-digital materials will obviate any initiative to create large collections of digital surrogates.

The Cornell libraries have tried both collection- and user-driven approaches to selection. In several instances, staff members have begun with expressed interests of faculty, say for teaching, and have developed digital collections based on those interests. In each case, however, library staffs have expanded their brief and have augmented faculty choices with related materials. A faculty member's inter-

ests are usually fairly circumscribed, and librarians select a good deal of additional materials on a topic, such as Renaissance art, to add depth to a selection. As a result, a selection of materials becomes a collection and has a wider scope of content. Research librarians are used to thinking of collections as being useful to the extent that they offer comprehensiveness or depth. Scholars, on the other hand, take comprehensiveness for granted and concentrate on making choices and discriminations among collection items in order to build a case for an interpretation. These two views of collections are complementary; however, when it comes to selection for digitization, they create the most difficult choices facing libraries in digitization programs. Selection is an "either/or" proposition. It seldom tolerates "both/and" solutions. Those historians who are working on Gutenberg-e projects sponsored by the American Historical Association are beginning to encounter the limitations that librarians live with every day. When faced with the opportunity not only to write their text for electronic distribution but also to present their sources through digital surrogates, the historians find themselves facing dilemmas familiar to digital collection builders everywhere. How much of the source material is enough to represent the base from which an argument was built? How can one select materials that give a sense of the scope of the original from which the scholar made his or her choices? And why is digitization of even a few core files so expensive?

Many of the scholar-designed projects may be coherent digital objects in themselves, but they would fail the librarian's test of comprehensiveness as a collection. Indeed, one could say that the value added by the scholar lies precisely in its selectivity. Some of those projects, most notably the Valley of the Shadow Civil War site, attempt to bring together materials that complement and enrich each other but do not try to comprehend the great universe of materials that could be considered complementary. These new digital collections are somewhat analogous to published anthologies of primary sources, carefully selected by an individual or an editorial team to serve heuristic purposes or to provide supporting evidence for an interpretation. Other projects driven by scholar selection, such as the "Fantastic" collection of witchcraft and French Revolutionary source materials at Cornell or the Women Writers' Resources Project at Emory University, do not claim to be comprehensive, but serve as pointers to the collection by presenting a representative sampling of it. Yet others, such as the Blake Archive or the APIS serve primarily to collocate items to form a new virtual collection that then serves as a new paradigm of critical edition.

General Collections. To date, very few libraries have digitized significant series of books and periodicals, whereas, as the ARL special collections survey shows, a great many libraries are digitizing their special collections. Several reasons for this selection strategy are commonly given, and others can be inferred.

There has been a preference to digitize visual resources over textual sources, in part because they work so well online and in part be-

cause visual resources do not require the additional expense of OCR or text encoding that add value to textual materials. (Creating metadata for visual resources that are not well indexed, however, often ends up being more expensive.) Printed sources do not require additional features, of course, but simple page images of non-rare texts do not provide the enhanced access that most researchers want from digital text. Nearly all selection criteria call for a specific additional functionality, such as browsing and searching, from text conversions.

A number of commercial interests are working with publishers or libraries to provide digitized versions of texts that have a potential market; one example is Early English Books Online (EEBO). For the core retrospective scholarly literature that is in high demand, there are commercial and nonprofit providers, such as Questia and JSTOR, ready to run the copyright gauntlet that libraries are ill equipped to handle efficiently. ArtSTOR, a production-scale digitization program initiated in the summer of 2001 and modeled on JSTOR, will develop a database of art and architecture images based on analysis of curricular needs for higher education. It is the copyright issue associated with these sources that has largely precluded individual institutions from digitizing their slide libraries and other visual resources and contributing them to a database that would build contextual mass.

In addition to these considerations, there is a sense that a library can help build institutional identity by digitizing materials that are unpublished or not commonly held. This can be important in encouraging alumni loyalty or in recruiting students. This assumption—that special collections build institutional identity and general collections do not—is actually challenged by the success of the Making of America (MOA) projects at Michigan and Cornell and by the texts encoded at UVA. These institutions have achieved considerable renown for their collections of monographs and periodicals. Yet it is reasonable to assume that such massive digitization programs are not easily replicated by institutions with smaller digital infrastructure. For those institutions, special collections allow a smaller-scale approach to developing a Web presence.

But, as the MOA projects highlight, two of the challenges faced by libraries mounting print publications are how much is too little and how much is perhaps too much. The sense that textual items need to exist in a significant or critical mass online stems in part from the fact that books and magazines do not have quite the same cultural frisson as do Jefferson holographs or Brady daguerreotypes. Few libraries are in the position to mount the kind of large-scale digitization projects that can result in a critical mass of text online, and for reasons to be discussed below, they do not enter into the kind of cooperative arrangement that is at the heart of MOA.

Special Collections. In 1995, when several academic libraries were working together to mount the text-based Americana in the Making of America project, LC inaugurated a digitization program, American Memory, based on its Americana special collections. The program was ambitious (they targeted five million images in five

years) and has been influential largely because of the extensive and easily adapted documentation that the library has mounted on its Web site and the well-publicized redistribution grants that it gave under its LC/Ameritech funding. The requirements for those grants were based on Library of Congress experience, and they have significantly influenced the requirements of other funding agencies, including the Institute for Museum and Library Services (IMLS). The only other library that has similar collecting policies and a similar governance and funding structure is the NYPL, and the digital program it plans to implement over the next few years bears remarkable resemblance to that of LC: both have an ambitious time frame, focus on special collections, and intend to make access to the general public as high a priority as service to scholars. They share, in other words, the same strategic view of digitization—one that is well in line with the realities of their roles as public institutions and with their audiences, collection strengths, and governance structures.

Many libraries that are not similar to LC or NYPL have also used this strategy. In the early stages of its digitization program, Indiana University reports that it used the LC/Ameritech Competition proposal outline to assess the merits of collections for digitization. This led to a canvass of the university's libraries for "their most significant collections, preferably ones in the public domain or with Indiana University-held copyrights." Then, with (special collection) candidates in hand, the library examined them for what it identified as the basic criteria: "the copyright status of the collection; its size; its popularity; its use; its physical condition; the formats included in the collection . . . and the existence of electronic finding aids" (Brancolini 2000).

Libraries that depend on outside funding—the great majority of libraries digitizing collections—often assert that it is easier to raise funds if they propose to digitize special collections because they are more interesting and have greater appeal to the funding agencies. This hypothesis is untested—although MOA, which comprises general collections, has received major grant funds, so perhaps the hypothesis has indeed been tested and proven invalid. Nevertheless, this notion of the funding bodies' predilection for special collections continues to be persuasive.

While academic libraries have many reasons for deciding to digitize special collections, the rationales of the two public institutions merit special consideration, in large part because they are so different from those of academic libraries. (Because of these differences, however, taking them as a model should be done with eyes wide open). The NYPL and LC base their selection decisions on their understanding that they are not libraries within a specific academic community, with faculty and students to set priorities. Rather, they serve a broad and often faceless community—the public. Their goal is to make available things that both scholars and a broader audience will find interesting. They also endeavor to make their collections accessible to those with modem connections and low bandwidth, often limiting factors for the delivery of cartographic and audio mate-

rials, among others. Because their primary audience is not academic, they have no curricular or educational demands to meet. They can focus exclusively on their mission as cultural institutions. Moreover, as libraries that have rich cultural heritage collections held in the public trust, they feel obligated to make those unique, rare, or fragile materials that do not circulate available to patrons who are unable to come to their reading rooms. Their strategic goal is cultural enrichment of the public.

None of the research libraries with comparable deep collections claims cultural enrichment of the public as an explicit goal. And yet, as de Stefano points out, there are academic libraries that are mounting special collections of broad public appeal that is not matched to curricular needs (de Stefano 2001, 67). She cautions that, "It is only a matter of time until the question emerges as to how long the parent institutions will be satisfied with supporting the costly conversion of their library's materials to improve access for narrowly defined audiences that may not even be their primary local constituents." This strategy may be supportable as long as the overhead of serving a secondary audience beyond the campus is low. But anxiety about this issue was repeatedly expressed in the conversations held during research for this paper.

3. Institutional Impacts

Libraries face several issues as a consequence of undertaking digital conversion. Some of these issues are constraints that may limit the potential of digitization to enhance research and teaching. They must be identified and explored to assess accurately the costs and benefits of digitization. Other issues may not be constraining as such, but may put new pressures on libraries to expend further resources. They, too, must be scrutinized and managed if digital programs are to be developed cost-effectively and with the greatest possible benefit to the collections and their users.

3.1. Treatment and Disposition of Source Materials

Selecting for scanning must include an assessment of the source's physical condition and readiness for the camera. For those items that are rare, unique, fragile, or otherwise of artifactual value, preparation for scanning usually demands the attention of conservators, if only to confirm that the item will not be harmed by the camera. Some type of prescanning treatment may also be required, from strengthening or mending paper to removing environmental soil. In the case of certain media, such as medieval manuscripts or daguerreotypes, the institution may need to call upon the services of a consultant or vendor with special expertise. This involves more time and resources (e.g., for writing contracts and checking items that return from a vendor) and may divert resources normally dedicated to other conservation work.

The disposition of scanned source materials that are not unique or rare is a challenging subject that most libraries are just beginning to confront. When digitization becomes an acceptable, if not preferred, alternative to microfilm for preservation reformatting and those items that can be networked are, what criteria will libraries use to decide what to keep on campus, what to send to remote storage, and what to discard? For materials that are rare or unique that question should not arise. But what about back journals that will be available from a database such as JSTOR, or American imprints that the University of Michigan and Cornell have scanned and made available without restrictions on their MOA site? The library community never reached a consensus about this issue for microfilming. The historical inattention to originals on the assumption that some institution would retain an original has led to avoidable material losses and to a serious public relations problem.

Recent trends in scholarship have forced libraries to reexamine old assumptions about the value of original serials and monographs—items that are not traditionally prized for their artifactual value (Council on Library and Information Resources 2001). It is important that research libraries find cost-effective means of preserving originals and making them readily accessible. This does not mean that all libraries should keep multiple copies of items that researchers prefer to use electronically. But 20 years from now, when many scholars may well prefer accessing these materials electronically over retrieving them from library shelves, will the library community have developed a collective strategy for preserving a defined number of originals for access purposes and reducing the redundancy of print collections? How will researchers who wish to have access to originals be able to find out where they are and how they can be viewed? Current plans to register digitized items in a nationally coordinated database include consideration of noting the location of a hard copy available for access, and the fate of originals is beginning to get the attention that researchers demand (Digital Library Federation 2001).

For certain very fragile media, such as lacquer sound discs or nitrate film, the original or master should seldom or never be used for access purposes. Service of those collections should always be done on reformatted (access) media. However, that is an expensive proposition for any library, and there is great resistance to push the costs of preservation transfer onto the user. Regrettably, a great number of recorded sound and moving image resources continue to be played back using the original. Until digitization is an affordable option for access to these media, their preservation will remain at very high risk.

3.2. Scalability

Digitizing either general or special collections presents challenges regarding size: How many items from any given collection will be sufficient to create added value? "Critical mass" is one selection cri-

terion that shows up in nearly all the written guidelines for selection and is commonly noted in conversation. The magic of critical mass, in theory, is that if you get enough related items up in a commonly searchable database, then you have created a collection that is richer in its digital instantiation than in its analog original. This is premised on the notion that the technology has a transformative power—that it can not only re-create a collection online but also give it new functionality, allow for new purposes, and ultimately create new audiences that make it available for novel queries. It does this by, for example, turning static pages of text or numbers into a database. Monographs are no longer limited by the linear layout of the bound volume or microfilm reader. By making texts searchable, librarians can create new resources from old ones and transform items that have had little or no use into something that receives hundreds or thousands of hits.

But how much is enough? A critical mass is enough to allow meaningful queries through curious juxtapositions and comparisons of phenomena, be it the occurrence of the word "chemise" in a run of Victorian novels or the U.S. Census returns from 1900. A large and comprehensive collection is valuable because it provides a context for interpretation. But in the digital realm, critical mass means something quite new and as yet ill-defined. The most salient example of this phenomenon is the MOA database at the University of Michigan, which contains thousands of nineteenth-century imprints at risk of embrittlement or already embrittled. Staff report that although the books themselves were seldom called from the stacks, the MOA database is heavily used, and not exclusively by students and teachers of Michigan. Members of the University of Michigan community use MOA most heavily, but among its largest users is the Oxford University Press, which mines the database for etymological and lexical research. Is this database heavily used because it is easily searched and the books were not? Because one can get access to it from any computer in any time zone, while the books were available to only a small number of credentialed users? Were the books of no research value when they languished in remote storage? And what is their research value now? It is hard to isolate a single factor that is decisive. Attempts to create equally valuable critical masses must address not only content but also searchability, ease of use, ranking on various search engines, and so forth.

In the case of general collections, or imprints, one must select enough that they, taken together, create a coherent corpus. In one case, time periods may set the parameters; in another, it might be genre or subject. In many ways, what makes a digitized collection of print materials singularly useful is the ability to search across titles and within subjects. The more items in the collection, the more serendipitous the searching. But as JSTOR has shown, incremental increase in the number of titles in the corpus is possible because they already exist within a meaningful context and are available through a nimble search and retrieval protocol.

Critical mass could more accurately be thought of as "contextual mass," that is, a variable quantity of materials that provides a context for evaluation and interpretation. In the analog realm, searching within a so-called critical mass has always been very labor-intensive. It has taken great human effort and patience to identify the relationships in and among items in a collection, and it has been possible only within collections that are physically located together. But once those items are online, in a form that is word-searchable, one has a mass that is accessible to machine searching, not the more arduous human researching. Theoretically, when many related collections exist together on the Web, they create a significantly more meaningful source than these same collections would if not linked electronically. In reality, achieving such a contextual mass across institutional boundaries will remain elusive as long as the collections are not interoperable.

For archival-type collections, which are not necessarily text based and are usually under looser bibliographical control than are published works, the amount of material needed to get a critical mass can defy the imagination, or at least challenge the budget. If a collection is too large to digitize—for example, a photo morgue or institutional records—staff may choose to digitize a portion that represents the strengths of the collection. But what portion? How much is enough? These are subjective decisions, and they are answered differently by different libraries. In the public libraries, with no faculty to provide advice, the decisions have been made by the curatorial staff. The Library of Congress has called in scholarly consultants and educational experts from time to time to aid in selection decisions, but the curatorial staff always makes the actual decisions. The NYPL relies on a curatorial staff that is expert in a number of fields. Like most cultural heritage institutions, it has long corporate experience in selecting for exhibitions. Curatorial staff in academic special collections libraries often have the opportunity to work with faculty or visiting researchers who collaborate in shaping a digital collection and even add descriptive and interpretive text to accompany items.

But many curators see digitizing anything less than a complete collection as "cherry picking," which results in a collection that does not support the research mission of the institution. Others are less severe and cheerfully admit that for most researchers, a little bit is better than nothing at all, and very few researchers mine any single collection to the depth that we are talking about. Those who do, they assert, would end up seeing the collection on site at some point in any event. These judgments are generalized from anecdotal experiences and are not based on objective data. When asked, for example, about how research techniques in special collections may be affected by digitization, some librarians said that research will be pursued by radically different strategies inside of a decade. Others think that research strategies for special collections materials will not change, even with the technology. The important thing, in their view, is not to get the resources online but to make tools for searching what is available in libraries readily accessible on the Web—tools such as finding

aids. The NYPL has secured money to do long-term studies of users of digitized special collections to gather information about use and to test assumptions. More needs to be done. A significant portion of grant-funded digitization, especially that supported by federal and state funds, should include some meaningful form of user analysis.

The California Digital Library (CDL) has inaugurated a project, called California Cultures, designed to make accessible a "critical mass of source materials to support research and teaching." Much of this documentation will reflect the social life, culture, and commerce of ethnic groups in California" (CDL 2000). The collection will comprise about 18,000 images. The California Digital Library sees collaboration as a key element in scalability. Because of funding and governance issues, the CDL believes that it must foster a sense of ownership and responsibility for these collections among creators statewide, locality by locality.

The role of scholars in selecting a defined set of contextually meaningful sources often works well for published items in certain disciplines. Agriculture and mathematics are examples where scholars have been able to come up with a list of so-called core literature that is amenable to comprehensive digitization. By way of contrast, curators may do a better job in selecting from special collections of unpublished materials—musical manuscripts, photo archives, personal papers—than scholars do. These are materials with which only curatorial staff members are sufficiently familiar to make selection decisions. While there are exceptions to this rule, the sheer quantity of materials from which to select often makes the involvement of scholars in all decisions impractical and hence unscalable.

Scholars tend to have a different concept of the critical mass than do librarians. Projects such as the Blake Archive and the Digital Scriptorium are built with the achievement of a critical mass for teaching and research in mind. Whereas a collection-driven, text-based program such as MOA can convert massive amounts of text and make it searchable, it can also put up materials without an interpretive framework. Other projects, such as APIS and, to a large degree, American Memory, invest time and money in creating interpretive frameworks and item-level descriptions that never existed when the items were analog, confined to the reading room, and served by knowledgeable staff. In many ways, this type of access is really a new form of publishing, not library service as it has been historically understood.

3.3. Intellectual Control and Data Management

The scarcity of cataloging or description that can be quickly and cheaply converted into metadata is often a decisive factor in excluding a collection from digitization. Given that creating metadata is more expensive than is the actual scanning, it is necessary to take advantage of existing metadata—that is, cataloging. Often, money to digitize comes with promises by library directors that they will put up several thousand—even million—images. This is a daunting

pledge. To mount five million images in five years, as LC pledged to do, has meant giving priority to large collections that already have extensive bibliographical controls. The NYPL is likewise giving selection preference to special collections that already have some form of cataloging that can be converted into metadata to meet production goals. In this way, expedience can theoretically be happily married to previous institutional investments. These libraries have put enormous resources into creating descriptions, exhibitions, finding aids, and published catalogs of prized institutional holdings. One can assume that a collection that has been exhibited or made the subject of a published illustrated catalog has demonstrated research and cultural value.

Some collections that are supported by endowments can make the transition to digital access more easily than can others, because funds may be available for this within the terms of the gift. The Wallach Division of Art, Prints, and Photographs, for example, at the NYPL will be put online as the Digital Wallach Gallery. There are a number of grant applications that not only build the cost of metadata creation into the digitization project but also appear to be driven in part by a long-standing desire on the part of a library finally to get certain special collections under bibliographical control.

It can be quite difficult, however, to harmonize descriptive practices that were prevalent 40 years ago with what is required today. The expansive bibliographical essays that once were standard for describing special collections need quite a bit of editing to make them into useful metadata. It is not simply a question of standards, which have always been problematic in special collections. The problem is that people research and read differently on the Web than they do when sitting with an illustrated catalog or finding aid at a reading desk. Descriptive practices need to be reconceptualized for presentation online. This reconceptualization process is several years off, since we have as yet no basis for understanding how people use special collections online.

For monographs and serials, genres for which the MARC record was originally devised and which is a standard well understood, retooling catalog records need not be complicated or expensive. For materials that are published but not primarily text based, such as photographs, posters, recorded speech, or musical interpretations, the MARC record has noted limitations and those tend to be accentuated in the online environment. Unpublished materials share this dichotomy of descriptive practice between textual and nontextual. For institutions that have chosen to put their special collections online, tough decisions must be made about how much information can be created in the most cost-effective way. In some cases, rekeying or OCR can be used to produce a searchable text in lieu of creating subject access. For handwritten documents, non-Roman scripts, and audio and visual resources, searching remains a problem.

The context for interpretation needs to be far more explicit in the online environment than in the analog realm. It is interesting to think about why that is so. Are librarians creating too much descriptive

material for online presentation of those collections that have successfully been served in reading rooms with no such level of description, or is it too little? Are librarians assuming that the level of sophistication or patience of the online user is far lower than that of the onsite researcher? It is commonly assumed that an online patron will not use a source, no matter how valuable, if it is accompanied by minimal-level description. This assumption may be well founded in principle, and it is certainly true that the deeper and more structured the description, the likelier it is that the item will be found through the various searching protocols most in use. But by removing research collections from the context in which they have traditionally been used—the reading room—one also removes the reference staff who can guide the patron through the maze of retrieval and advise about related sources. Materials that the public has had little experience using are now readily available online. If patrons lack certain research skills, the resources will remain inaccessible to them.

The ease of finding digitized items on library Web sites varies. A few sites are constructed in a way that makes finding digitized collections almost impossible for people who do not already know they exist. Other sites have integrated the surrogates into the online catalog and on OCLC or RLIN or both. Those DLF members whose primary purpose in digitization is to increase access to special collections and rare items have expressed willingness to expose the metadata for these collections to a harvester using a technical framework established for the Open Archives Initiative.

3.4. Coordinated Collection Development

The idea of coordinated collection development of digital objects is a powerful one. It motivated the Berkeley, Michigan, and Cornell libraries to work together to mount several collections from their own holdings that could be termed part of one, the Making of America. Given the resources that must be dedicated to creating digital collections and the resources it takes to build the infrastructure that allows access to them, it would seem that the only way to build truly scalable collections is through some cooperative effort.

But for all the talk of building federated collections that will aggregate into a digital library with depth and breadth—that is, critical mass—the principle of "states' rights" remains the standard. Each institution decides on its own what to digitize, and usually does so with little or no consultation with other libraries. There are funding sources that require collaboration in some circumstances—the Library of Congress's Ameritech grant is an example—but the extent of collaboration usually has to do with using the same standards for scanning and, sometimes, description. Selection is not truly collaborative; it could more properly be characterized as "harmonized thematically." Institutions usually make decisions based on their particular needs rather than on community consensus about priorities.

How do we know that we are not duplicating efforts in digitization, even when the content comes from special collections? Two

sheet music projects, the Levy Sheet Music Collection at Johns Hopkins University and the Historical American Sheet Music Collections at Duke University, may or may not have significant overlap, for example. There may be sound reasons to scan each collection in full, even if there is overlap. But we are not able to make that kind of decision at present without a comprehensive database about such projects. Some will argue that duplication is good and that dreams of standardization are premature. Nonetheless, while duplication may have some benefits at these early stages of research and development, we are unable to take advantage of these benefits unless we know what others are doing. A registry that would make core pieces of information available, not only about content but also about technical specifications, disposition of originals, and cataloging preferences, is a critical component of infrastructure that would allow each institution to make informed decisions on a series of matters (Digital Library Federation 2001).

3.5. Funding

Some library staff worry that institutional concerns, such as fund raising, public relations, and special projects, divert too many resources from more academically defensible projects or from the core mission of the library. When asked to take on projects conceived to advance the home institution's mission, rather than the research and teaching mission of the library per se, most library administrators seem willing to accept this "good citizen" role, and some use it to the library's distinct advantage. Even a vanity project, if managed properly, will bring money into the library for digitization and provide the kind of training and hands-on experience that is necessary to develop digital library infrastructure and expertise. The key to building on such a project is to be sure that *all* the library's costs, not only scanning but also creating metadata, migrating files, and so forth, are covered. Such projects, done willingly and well, usually enhance the status of the library within the community and seldom do long-term harm. The issue bodes ill only when libraries deliberately seek funding for things that are not core to their mission or when staff and management are diverted to support low-priority projects. Looked at from this perspective, outreach can properly be considered part of the library's mission.

Academic libraries such as those at Virginia and Cornell, which are funded by a mix of private and public monies, are liable to face pressures to serve not only research and teaching needs but also state and regional interests. These need not be mutually exclusive, and even Harvard University has demonstrated its good citizenship by contributing items of interest to Cantabridgeans on its publicly accessible Web site. The key lies in achieving a balance and, if possible, a synergy between the two.

For public institutions, digital programs offer a new and unique way to serve collections to taxpayers. For example, online distribution is the only way LC can provide access to its holdings in all con-

gressional districts. For the primary funders and governors, that is, members of the U.S. Congress, who have built and sustained the library on behalf of their constituents, this rationale is compelling. Similarly, while the NYPL is not fully supported by public funds, its choices are nonetheless strongly influenced by the priorities of state and local governments.

Academic libraries with dual funding streams—private and public—are most vulnerable when the state in which they are located expresses some expectation that the university library will mount materials from its collections that are aimed at the kindergarten through twelfth-grade user group. There is much talk of how access to primary source materials held in research institutions can transform K–12 education. This hypothesis needs to be tested. The fact that much grant funding is tied to K–12 interests has made it necessary for universities to try to shape research-level materials into a K–12 mold to secure funding, or at least to mask a research collection as one that is also suited for younger audiences.

There is no doubt that public institutions are seen as holding a promise to improve the quality of civic life if they provide greater access to their holdings. The fact that the NYPL and LC have been so successful in securing funds from private citizens is a clear indication of the public esteem that these institutions enjoy and the desire to "get the treasures out." This level of philanthropy would be unthinkable in any other country. While these libraries may be accused of pandering to donors on occasion or of not paying enough attention to the academic community by digitizing materials that are not in demand first and foremost by scholars, the fact is that they are public libraries. Unlike the libraries in state universities, they are not designed to serve exclusively, or even primarily, the scholarly community. This obligation to serve the public, however, does not skew selection for digitization as drastically as some assert. Donors may express an interest in a particular type of material, but in the end, they choose from a set of candidate collections that have been proposed by curatorial divisions and vetted by preservation and digital library staff for technical fitness. In terms both of process and result, these candidate collections differ little from their private academic counterparts.

In both public and private libraries, some curators who are active in special collection development advocate for digitization because they see it as a way to induce further donations. For them, the promise of access is a useful collection-development tool because digital access advertises what the library collects and demonstrates a commitment to access.

3.6. Preservation

Much less has been written about how to plan for the access to and preservation of digitally reformatted collections over time than about how to select materials for digitization. This is partly because we know nothing certain about maintaining digital assets over the long

haul. We have learned a great deal as a result of failed or deeply flawed efforts—those of the "We'll never do that again!" variety told of some projects to reformat information onto CDs, for example—but such lessons tend to be only informally communicated. Exceptions include the University of Michigan, where one library that has a clear view of what role digitization plays—that of collection management and preservation—has developed and published preservation policies that support those goals. The CDL is also an exception, perhaps because, as a central repository, establishing standards and best practices to which its contributors must adhere is paramount to building confidence as well as collections. Harvard University has published information about its plans for a digital repository, and Cornell has adopted policies for "perpetual care" of digital assets. The Library of Congress has also put online much about its planned audiovisual repository in Culpeper, and has announced a plan to develop a national digital preservation infrastructure (LC 2001). General information about the preservation of digital files can be found on the Web sites of the Cedars Project, CLIR, and Preserving Access to Digital Information (PADI).

Nearly every library declares its intention to preserve the digital surrogates that it creates. The Library of Congress has also pledged to preserve those surrogates created by other libraries under the auspices of the National Digital Library Program (Arms 2000). In reality, however, many libraries have created digital surrogates for access purposes and have no strategic interest in maintaining them as carefully as they would have if they had created those files to serve as replacements. Nonetheless, at this point libraries are uncomfortable admitting that they have a limited commitment to many of their surrogates, should push come to shove. Those who are creating surrogates for access purposes alone still declare an interest in maintaining those surrogates as long as they can because the original investment in the creation of digital files has created something of enormous value to their patrons. Moreover, the cost of having to recreate those surrogates and the physical stress it might impose on the source materials argue for maintaining those files as long as possible.

The mechanism for long-term management of digital surrogates is theoretically no different from that for management of born-digital assets. While refreshment and migration of digital collections have occurred in many libraries, the protocols and policies for preservation are clearly still under development. Many libraries have been sensitized to the fact that loss can be simple and catastrophic, beginning with the wrong choice of (proprietary) hardware, software, or medium on which to encode information and ending with negligent management of metadata. The Y2K threat that libraries faced in 1999 has led to systemic improvements in many cases. Not only did institutions become aware of how deleterious it is to allow different software to proliferate but they also developed disaster-preparedness plans. Many received funds for infrastructure upgrades that might have been awarded much later, or not at all, were it not for the sense of urgency that the coming of Y2K provided.

Libraries are anticipating the day when they must develop strategies for handling digital objects that faculty are creating without the involvement of the library. These are the often-elaborate constructions done by individual scholars or groups of collaborators that the library hears about only after the critical choices of hardware, software, and metadata have been made, often by people wholly unaware of the problems of long-term access to digital media. An increasing number of library managers have expressed concern about the digitized materials created by faculty that are "more than a Web site" yet less, often far less, than what the library would choose to accession and preserve. While libraries acknowledge that this is a growing problem, none has been forced to do much about it yet, and thoughts about how to deal with faculty projects are just now evolving. Predictably, those that are collection-driven in approach are working to build a system for selecting what the library wishes to accession to its permanent collections. Cornell is developing criteria that individuals must work to if they expect that the library will provide "perpetual care" (Cornell 2001). CDL already has such guidelines in place. Michigan has a well-articulated preservation policy, one that is detailed enough to support the university's vision of digital reformatting as a reliable long-term solution to the brittle book problem.

3.7. Support of Users

There is little understanding of how research library patrons use what has been created for them. Most libraries recognize that the collections they now offer online require a different type of user support than that which they have traditionally given visitors to their reading rooms. In many cases, user support has been developed for "digital collections" or "digital resources," terms that almost invariably denote born-digital (licensed) materials. The Library of Congress, which targets a K–12 audience, has three reference librarians for its National Digital Library Program Learning Center. As a rule, however, libraries have not been reallocating staff to deal specifically with digitized collections. Hit rates and analyses of Web transactions have yielded a great amount of quantitative data about access to digital surrogates, and those data have been mined for a number of internal purposes, from "demonstrating" how popular sites are to making general statements about how users are dialing in and from where. Qualitative analysis is harder to derive from these raw data, and there have been few in-depth analyses of how patrons are reacting to the added functionality and convenience of materials now online. Libraries have been keeping careful track of gate counts, for example, but when these counts go up or go down, what conclusion can one draw about the effect of online resources on use of on-site resources?

One exception to this rule is the journal archiving service, JSTOR, which rigorously tracks the use of its resources. JSTOR analyzes its users' behavior because it needs to recover costs and hence must stay

closely attuned to demand, within the constraints of copyright issues. Looking ahead, close analysis of how researchers use specific online resources, especially how they do or do not contribute to the productivity of faculty and students, will be a prime interest of libraries. Further work must begin to complement this work with analysis of free Web sources mounted by libraries.

Most libraries report having classes and other instructional options available for students and faculty. Some librarians report that instruction is not really necessary for undergraduates, who are quite used to looking first online, but that general orientations to library collections are needed more than ever.

Good Web site design, that is, creating sites that are easily found, easily navigated, and readily comprehensible, is an often overlooked aspect of access. Within the first-generation libraries, there is an astonishing variety in the quality of Web site design. Even for fairly sophisticated users, finding a library's digital collections can involve going through a half-dozen screens. Having a professional design team that keeps a site up-to-date and constantly reviews it for improvements does cost money, but considering that the Web site is the front door to the collections, it would seem penny wise and pound foolish to ignore design and marketing.

4. Conclusions and Recommendations

Digitization, like other collection-development strategies, works to the extent that it supports the mission of an institution. Without a clear sense of how digital projects may fit an institution's mission, it is difficult to build a strategy for sustaining them. It may still be too early for libraries to be thinking far into the future as far as digitization is concerned. This is a period of experimentation, of building skills and experience in staff, and of tracking one's own progress and that of others. In some libraries, the staff feels driven by the need to have some digital projects, whether or not library leaders have made clear what purpose such projects will ultimately serve. For the time being, modest strategies may be the best route. Simply declaring a project or series of projects as experiments and stating what goals can be achieved through those experiments will help begin to free staff for creative and reflective activity. Certain digitization approaches work chiefly for large libraries and make little sense for small ones; others work for large public libraries but not large academic ones. Not all strategies are scalable, and any digital projects that exist in splendid isolation from other parts of the home institution or other libraries risk turning into a waste of resources.

4.1. Costs

A useful way of looking at the strategic value of digitized collections is to ask pointed questions that reflect perceived costs and benefits, such as the following: What would happen if these programs re-

ceived a dedicated allocation? Would the money for creating digital surrogates come from the acquisitions budget? From preservation? If the program were supported through a separate line, from which pocket of money would these funds be reallocated? Should fees be charged for access, at least in some cases?

Michigan, which says that it left project-based digitization three years ago, has ongoing budget support for the staffing, equipment, and other infrastructure services such as servers and software resources for digital conversion production. The university anticipates that future projects will require additional funds to support special needs outside the core capacities or to handle a larger volume of work. Harvard's Digital Library Initiative is supported by internal funds, although its purpose is not collection development primarily but building infrastructure.

At most libraries, digitization costs are covered by external funds, and the projects developed appeal to the intended funding source, be it a federal agency with stringent grant conditions; a private foundation that has a heuristic interest in projects; or donors and alumni, who usually contribute to the institution for eleemosynary purposes and often do so out of dedication to the institution and its mission per se. When asked about their priorities for selection, many respondents remarked wryly that they digitize what they can get money to do, implying and even sometimes stating directly that their choices were skewed by donors' priorities and did not serve pure scholarship or other core missions of their institutions. However, what selectors, curators, and bibliographers think to be of highest value will also often differ from what the administration identifies, because they have differing views of where scholarship is moving, how sophisticated the users are, and what is of lasting import.

Some librarians expressed great concern about the fact that, as long as libraries are competing for outside funds to digitize, they will be stuck in the entrepreneurial phase in which collection development is driven by strong personalities—those who are willing to compete for funds—and that some parts of the library's collections will go untapped simply because the subject specialist in that area is not the "entrepreneurial" type. Others express more serious concerns about the fate of non-English language materials and even greater anxiety about the neglect of non-Roman collections.

Concerns about the changing role of library staff, especially of bibliographers, come up with increasing frequency. Staff become increasingly diverted from traditional collection-development duties to spend more time selecting for digitization—what might be called "reselecting." This is bound to have some effect on current collection development of traditional materials. A topic far more widely discussed is where to find the skill sets that are needed for digital library development. If libraries cannot afford to hire persons with such skills—as increasingly they cannot—how are they to go about developing them internally without robbing Peter to pay Paul? The same is true for preservation staff, who see themselves and their

funding diverted from preserving deteriorating collections to creating digital versions of materials that are often not at imminent risk of deterioration.

A number of special collections librarians have noted that, as items from the collections are digitized and given visibility, more people have become interested in the items and in related, not-yet-digitized materials. The numbers of on-site users and of phone, letter, and e-mail inquiries have risen. This places increased physical stress on original materials. This, in turn, increases the workload on staff, especially on preservation and reference staff.

Reliable and meaningful cost data about digitization are rare and not often useful in comparative contexts. Costing out the elements of digitizing means beginning with selection and going to physical preparation, cataloging, physical capture, creation of metadata, mounting and managing files, designing and maintaining the site, providing additional user services, and going through to implementing a long-term preservation strategy. Virtually every step in digitization involves human intervention and skill, and these costs, unlike those of storage, for example, are unlikely to go down.

4.2. Benefits

In an exercise not yet seen in the United States, Oxford University libraries recently looked at their experiences with digital conversion in an attempt to identify what benefits it had brought to the library and its patrons (Oxford 1999). Curiously, one of the chief benefits cited was to curators, who learned an enormous amount about their own collections and those of other colleges. This seems an expensive way to break down barriers between departments and other library units or to make curators more familiar with their collections. But the report is alluding to one way in which the very role of bibliographer and curator is changing as a result of putting collections online. This is similar to reports from various campuses across the United States, which state that digital projects are crucial for staff training and developing expertise. The matter of collection expertise, however, is not often mentioned; instead, reports about benefits focus on the other aspect of so-called e-curatorship, in which staff members develop technical and editorial or interpretive expertise. This does not always benefit the library or university in the end, because expert staff are difficult to retain and many managers complain that expenses incurred in training a staff member served to benefit that individual's next employer.

Benefits to users are also commonly cited, though as noted, systematically gathered evidence of user satisfaction is rare. Managers also cite benefits to the collections through creation of surrogates that protect original materials while increasing access to the content.

4.3 Looking Ahead

If it is to be sustainable over time, digitization must clearly be an integral part of the core mission work of the library. Whereas the majority of research libraries engaged in digitization have been able to raise external funds for conversion, they all recognize the hazards of becoming too dependent on such funds. There is no such thing as a "free" building. Even if a donor were to pay for all aspects of the construction, from land acquisition to furnishing, at some point maintenance costs will become the responsibility of the home institution, and the building must meet minimum criteria for support.

The same holds true of digitized collections. Over the next few years, some libraries that have done digital projects will essentially phase them out; others will reduce this activity to the exception rather than the rule. Still others, committed to large-scale digital projects, either as a part of collection management or as a commitment to extending access, will continue and will begin to address the tough questions of finding internal funds or developing fee-based services to support conversion, maintenance, and service. Whatever the scenario, libraries will need to articulate how digitization serves their core mission, be it outreach, preservation, access, or something else. If libraries continue to focus primarily on digitization of their special collections for access to primary sources (as opposed to discrete exhibitions for outreach purposes), they must act now to start building common infrastructures, such as a registry to obviate duplication and a commitment to rigorous cooperative collecting. One of the great potentials of the technology is to reunite collections that are dispersed and to create new meta collections (such as Blake's works) and allow several formats to interact. These are rather modest goals best achieved in well-defined projects, and a far cry from the rather common decisions now made to put up a large, interesting collection of archival or special collection materials. Without appropriate tools to search across collections and architectures that support interoperability, such collections will prove inconvenient to find and laborious to use.

Looking back to the proceedings of a symposium on selection for digitization sponsored by the Research Libraries Group in 1995, we are confronted by similar concerns (RLG 1996). Among many factors identified at the meeting that remain valid today are the need for more focused selection of special collections, either with others to build a large cohesive body of materials or designed and scaled for classroom use; the need to engage collection development librarians in selection for reformatting; the need to develop technical specifications for a variety of things, from scanning to metadata creation, that can be widely embraced by institutions of different size and mission; the need to decide how to manage storage of surrogates and for how long; the need to find outside funds to digitize and not be constrained by that when selecting for conversion; and the need to identify how digitization would add value to the source material, from uniting disparate collections to making their contents more searchable.

A principal barrier to developing freely accessible digital collections of enduring value is the lack of sustainable economic models for digitizing collections, especially special collections. Each digitization program requires substantial investments of numerous kinds: processing and preserving source collections; creating interoperable systems for increased ease of access; copyright research and ongoing rights management models; and coordinated selection to increase the intellectual value of collections. On top of that are infrastructure services that need to be available for small or midsize institutions to keep the digital divide from deepening in the research community: digitizing and archiving services; a registry that would provide information about what others are doing; and accepted practices for the capture and creation of metadata that can be used in good faith. The benchmarks for good collection development—rigorous selection, accurate description, fitness-for-use, authenticity, and intellectual or cultural significance—do not change in the digital realm. It may be harder to know for certain how to build those collections online than in hard copy, but the core values of collections have not changed.

4.4. Recommendations

While it is too early to set standards or even memorialize best practices for selection of collections for digital reformatting, libraries, funders, and research, development, and policy groups can take actions that will minimize the chances that current investments will be wasted. These include taking deceptively simple steps in the course of decision-making, assessing progress and change, tracking expenditures across all phases of activity in a consistent manner, and engaging the widest possible group of people, from technologists to scholars, in selection, assessment, and planning. The following recommendations come from a variety of sources in the field. All are based on experience and reflection about what has worked to date, and what next steps must be taken to advance the building of sustainable and valuable digital collections.

Recommendations for Institutions

- Be clear about the purpose of the project—for example, creation of preservation surrogates, outreach, or curricular development.
- Begin with a clear understanding about whether or not the library will maintain the surrogates and for how long.
- Develop clear protocols for selection decisions.
- Develop cost projections for all aspects of digital conversion, from selection to refreshing of files.
- Either focus on materials that are already organized or target materials not under intellectual control; include funding for processing and description in the project budget.

- Work with faculty and scholars to develop controlled vocabularies in those fields lacking them, to aid in creating metadata.

- Clarify the target audience for a collection: do not attempt to be all things to all audiences.

- Secure funding for and conduct user assessments of digitized collections and make the information from these assessments available to the library community.

- Consider making metadata available for harvesting through the Open Archives Initiative protocol.

- Design a Web site that is graphically clear and easy to navigate; update it frequently as appropriate.

- Develop a policy for the disposition of reformatted materials.

Recommendations for Research Agenda and Agenda for Consortia and Funding Agencies

- To build coherent collections, extend editorial board methodology, used for agriculture and mathematics, to select core literature in other disciplines.

- Encourage funding agencies to support good practice by requiring minimal standards of capture, use of nonproprietary software, planning for maintenance of digital surrogates, and so forth.

- Urge donors and consortia to engender partnerships for cooperative digitization of both general and special collections.

- Examine the expectations of academic libraries' service to the broad public: what are the roles of municipal, state, and federally supported institutions for public access and those of private institutions?

- Conduct analysis of how digitized collections, both special and general, are used and by whom.

- Foster the development of a registry of digitized collections that includes capture information, metadata standards, and disposition of original source materials.

- Foster the development of service bureaus that would offer scanning and archiving services.

- Foster the development of shared print repositories.

5. References

Web addresses were valid as of September 6, 2001.

Arms, Caroline R. 2000. Keeping Memory Alive: Practices for Preserving Digital Content at the National Digital Library Program of the Library of Congress. *RLG DigiNews* 4(3). Available at http://www.rlg.org/preserv/diginews/diginews4-3.html.

Brancolini, Kristine R. 2000. Selecting Research Collections for Digitization: Applying the Harvard Model. *Library Trends* 48 (Spring): 783-798.

California Digital Library. 2000. Library of Congress Grant: California Cultures. Executive Summary. Oakland, Calif.: California Digital Library.

Cornell University. 2001. Report of the Digital Preservation Working Group on Establishing a Central Depository for Preserving Digital Image Collections (version 1.0). Available at http://www.library.cornell.edu/imls/image%20deposit%20guidelines.pdf.

Council on Library and Information Resources. 2001. *The Evidence in Hand: the Report of the Task Force on the Artifact in Library Collections* (draft). Available at http://www.clir.org/activities/details/artifact-docs.html.

Digital Library Federation. 2001. Registry of Digitized Books and Serial Publications. Available at http://www.diglib.org/collections/reg/reg.htm.

Digital Library Federation and Research Libraries Group. 2000. Guides to Quality in Visual Resource Imaging. Available at http://www.rlg.org/visguides/.

de Stefano, Paula. 2001. Selection for Digital Conversion in Academic Libraries. *College & Research Libraries* 62 (January):58-69.

Flecker, Dale. 2000. Harvard's Digital Library Initiative: Building a First Generation Digital Library Infrastructure. *D-Lib Magazine* 6(11). Available at http://www.dlib.org/dlib/november00/flecker/11flecker.html.

Gertz, Janet. 2000. Selection for Preservation in the Digital Age (Overview). *Library Resources & Technical Services* 44(April): 97-104.

Hazen, D., J. Horrell, and J. Merrill-Oldham, 1998. *Selecting Research Collections for Digitization*. Washington, DC: Council on Library and Information Resources. Available at http://www.clir.org/pubs/reports/hazen/pub74.html.

Kenney, Anne R., and Oya Y. Rieger, 2000. *Moving Theory into Practice: Digital Imaging for Libraries and Archives*. Mountain View, Calif.: Research Libraries Group. Available at http://www.rlg.org/preserv/mtip2000.html.

Library of Congress National Digital Library Program. 1997. NDLP Project Planning Checklist. Available at http://memory.loc.gov/ammem/techdocs/prjplan.html.

Library of Congress. 2001. "Library to Lead National Effort to Develop Digital Information Infrastructure and Preservation Program." Press release. January 12, 2001. Available at http://www.loc.gov/today/pr/2001/01-006.html.

Panitch, Judy. 2001. *Special Collections in ARL Libraries: Results of the 1998 Survey Sponsored by the ARL Research Committee*. Washington, D.C.: Association of Research Libraries.

Oxford Scoping Project. 1999. Scoping the Future of Oxford's Digital Collections. Available at http://www.bodley.ox.ac.uk/scoping/.

Research Libraries Group. 1996. *Selecting Library and Archive Collections for Digital Reformatting*. Mountain View, Calif.: Research Libraries Group.

Sitts, Maxine K., ed. 2000. *Handbook for Digital Projects: A Management Tool for Preservation & Access*. Andover, Mass.: Northeast Document Conservation Center (NEDCC). Available at http://www.nedcc.org/digital/dighome.htm.

Smith, Abby. 1999. *Why Digitize?* Washington, D.C.: Council on Library and Information Resources. Available at http://www.clir.org/pubs/reports/pub80-smith/pub80.html.

University of Michigan. 1999. Principles and Considerations for University of Michigan Library Subject Specialists. Guidelines for Selection in Digital Projects (Feb. 23 update). Available at http://www.umdl.umich.edu/policies/digitpolicyfinal.html.

Web sites

Advanced Papyrological Information System (APIS). Available at http://www.columbia.edu/cu/lweb/projects/digital/apis/index.html.

California Digital Library. Available at http://www.cdlib.org/.

Cedars Project. Available at http://www.leeds.ac.uk/cedars/.

Council on Library and Information Resources. Available at www.clir.org.

Digital Scriptorium. Available at http://sunsite.berkeley.edu/Scriptorium/.

Early English Books Online. Available at http://www.lib.umich.edu/eebo/.

Electronic Text Center. Available at http://etext.lib.virginia.edu/uvaonline.html.

Emory Women Writers. Available at http://chaucer.library.emory.edu/wwrp/index.html.

Garden and Forest. Available at http://lcweb.loc.gov/preserv/prd/gardfor/gfhome.html.

Gutenberg-e. Available at http://www.theaha.org/prizes/gutenberg/.

Hoagy Carmichael. Available at http://www.dlib.indiana.edu/collections/hoagy/.

Historic American Sheet Music Project. Available at http://scriptorium.lib.duke.edu/sheetmusic/.

Journal Storage (JSTOR). Available at http://www.JSTOR.org/.

Lester S. Levy Collection of Sheet Music. Available at http://levysheetmusic.mse.jhu.edu/.

Making of America. Available at
http://sunsite.berkeley.edu/MOA2/,
http://moa.umdl.umich.edu/, and
http://cdl.library.cornell.edu/moa/.

Mark Twain in His Times. Available at http://etext.lib.virginia.edu/railton/index2.html.

OAC (Online Archive of California). Available at http://www.oac.cdlib.org/.

Open Archives Initiative Protocol. Available at http://www.openarchives.org/OAI/openarchivesprotocol.htm.

Preserving Access to Digital Information (PADI). Available at http://www.nla.gov.au/padi/.

Questia. Available at http://www.Questia.com/Index.jsp.

Saganet. Available at http://saga.library.cornell.edu/saganet/apb2/?MIval=/StartPage.html (restricted site).

Sam Nunn Papers. Available at http://sage.library.emory.edu/collection-0800.html.

Scenery Collection. Available at http://digital.lib.umn.edu/scenery/.

The Complete Writings and Pictures of Dante Gabriel Rossetti. Available at http://jefferson.village.virginia.edu/rossetti/.

The Valley of the Shadow. Available at http://jefferson.village.virginia.edu/vshadow2/.

University of Virginia Library of Tomorrow. Available at http://staff.lib.virginia.edu/LofT/.

William Blake Archive. Available at http://www.blakearchive.org/.

BEST COPY AVAILABLE

COUNCIL ON LIBRARY AND INFORMATION RESOURCES

1755 Massachusetts Avenue, NW, Suite 500, Washington, DC 20036-2124
Telephone 202-939-4750 • Fax 202-939-4765 • Internet address: info@clir.org • Web: www.clir.org